

ABSTRACT

Systems and methods for word sense disambiguation, including discerning one or more senses or occurrences, distinguishing between senses or occurrences, and determining a meaning for a sense or occurrence of a subject term. In a collection of documents containing terms and a reference collection containing at least one meaning associated with a term, the method includes forming a vector space representation of terms and documents. In some embodiments, the vector space is a latent semantic index vector space. In some embodiments, occurrences are clustered to discern or distinguish a sense of a term. In preferred embodiments, meaning of a sense or occurrence is assigned based on either correlation with an external reference source, or proximity to a reference source that has been indexed into the space.